Phys. Med. Biol., 1987, Vol. 32, No 11, 1481-1494. Printed in the UK

# Bayesian image processing of data from constrained source distributions—fuzzy pattern constraints

Z Liang and H Hart

Department of Physics, City College of New York, NY 10031, USA
and
Department of Nuclear Medicine, Montefiore Medical Center and Albert Einstein College of Medicine, Bronx, NY 10467, USA

Received 5 November 1986, in final form 8 May 1987

**Abstract.** *A priori* probability density functions characterising patterns which are imprecise spatially and with regard to amplitude (fuzzy patterns) and which are anticipated to be present in a radioisotopic source field were developed for use in Bayesian image processing (BIP). Corresponding iterative imaging algorithms were derived using the expectation maximisation (EM) technique of Dempster *et al.* BIP and standard non-BIP algorithms were applied to computer generated and experimental radioisotope phantom imaging data. Improved results were obtained with BIP.

## 1. Introduction

In certain imaging situations, *a priori* information about the source distribution may exist and can be characterised in terms of a probability density function $P(\Phi)$ of the source elements $\{\phi_j\} = \Phi(j = 1, 2, \ldots, \alpha)$. In radioisotopic imaging applications, $P_j(\phi_j)$ can be the *a priori* probability distribution for photon emission from voxel $j$ per unit time at time $t = 0$. For example, in cardiac imaging studies it is possible to determine the radioisotope concentration of the blood filling the cardiac chambers as a function of time by peripheral measurement and to estimate the average radioisotope concentration of surrounding heart tissues. In brain imaging, hepatic imaging and in particular in thyroid imaging studies, probable configurations of cold or hot lesions may be characterisible in terms of radioisotope concentration and anticipated spatial relations between the cold or hot lesions (i.e. patterns). Lesions detection can then be viewed ideally as a search for the anticipated patterns superimposed upon a relatively uniform background†. For such cases, a Bayesian image processing (BIP) formalism which incorporates both *a priori* amplitude (strengths of pattern elements) and spatial information on source distribution and the statistical character of measured data (or projections) has been developed (Hart and Liang 1987). If the *a priori* amplitude and spatial information is probabilistic, the anticipated patterns are of course less well specified (more realistically) and can be classified as fuzzy.

In this paper, a BIP formalism for an *a priori* fuzzy pattern source distribution reflecting the Poisson nature of the measured data is represented and the effects of variations in the anticipated amplitudes and spatial relations of the fuzzy pattern elements are studied. Quantitative comparison of the BIP algorithm and the standard maximum likelihood non-Bayesian algorithm (Shepp and Vardi 1982, Lange and Carson 1984) indicates substantial improvement with the BIP algorithm.

† If the background is completely non-uniform and completely unpredictable then, of course, no image processing approach is likely to be feasible.

0031-9155/87/111481+14$02.50  

## 2. Probability density functions of fuzzy pattern and measured data

Consider first an ideally uniform background of radioisotope concentration (strength) $\phi_b^e$. The probability density function of this source distribution is simply a delta function. Assuming a fixed normal distribution about the average background value of $\phi_b^e$ for each voxel, the delta function becomes (operationally defining the term 'relatively uniform background'):

$$P(\Phi) = \prod_j^{\alpha} C_b \exp\left( -\frac{(\phi_j - \phi_b^e)^2}{2\sigma_b^2} \right) \tag{1}$$

where $C_b$ is the normalisation constant and $\sigma_b$ the standard deviation characterising the strength variation about the estimated $\phi_b^e$ (i.e. the uncertainty in the anticipated background amplitudes), and $\alpha$ is the number of voxels. For very small values of $\phi_b^e$, the normal distribution should be replaced by a Poisson distribution.

If on the average a single spot with estimated strength $\phi_1^e - \phi_b^e$ is superimposed on the background, voxel $j$ has either the value $\phi_1^e$ or $\phi_b^e$ with different probabilities, $1/\alpha$ or $(\alpha - 1)/\alpha$ (where $\alpha$ is the number of voxels), then function (1) becomes:

$$P(\Phi) = \prod_j^{\alpha} \left[ \frac{1}{\alpha} C_1 \exp\left( -\frac{(\phi_j - \phi_1^e)^2}{2\sigma_1^2} \right) + \frac{(\alpha - 1)}{\alpha} C_b \exp\left( -\frac{(\phi_j - \phi_b^e)^2}{2\sigma_b^2} \right) \right] \tag{2}$$

where $C_1$ and $\sigma_1$ are analogous to the $C_b$ and $\sigma_b$ respectively.

If it is anticipated that (as described in detail in appendix 1) a single pattern of two spots with estimated strengths of $\phi_1^e$ and $\phi_2^e$ respectively and separated by $l_1$ voxels is present somewhere in the source field, function (2) then has the qualitative form†:

$$P(\Phi) = \prod_j^{\alpha} \left\{ \frac{2}{\alpha} \frac{C_1 C_2}{2} \left[ \exp\left( -\frac{(\phi_{j-l_1} - \phi_1^e)^2}{2\sigma_1^2} - \frac{(\phi_j - \phi_2^e)^2}{2\sigma_2^2} \right) + \exp\left( -\frac{(\phi_j - \phi_1^e)^2}{2\sigma_1^2} - \frac{(\phi_{j+l_1} - \phi_2^e)^2}{2\sigma_2^2} \right) \right] + \frac{\alpha - 2}{\alpha} C_b \exp\left( -\frac{(\phi_j - \phi_b^e)^2}{2\sigma_b^2} \right) \right\}$$

or

$$P(\Phi) = \prod_j^{\alpha} \left\{ \frac{2}{\alpha} \frac{1}{Q} \sum_q \left[ \prod_{s=1}^{\beta=2} C_s \exp\left( -\frac{(\phi_{j+l_{qs}} - \phi_s^e)^2}{2\sigma_s^2} \right) \right] + \frac{(\alpha - 2)}{\alpha} C_b \exp\left( -\frac{(\phi_j - \phi_b^e)^2}{2\sigma_b^2} \right) \right\} \tag{3}$$

where the index $q$ covers the $Q$ possible pattern configurations and $\beta$ is the number of elements (or spots) in the pattern. Note that for a one-dimensional system the number of possible pattern configurations including voxel $j$ as a component in the pattern is $Q = \beta$. Thus:

$$l_{11} = -l_1 \qquad l_{12} = 0 \qquad l_{21} = 0 \qquad l_{22} = +l_1 .$$

Since the *a priori* value of $l_1$ may not be exact, a weighting function $W(l)$ is now introduced to specify the variation of $l$ around the estimated $l_1$ (i.e. reflecting the uncertainty in the anticipated spatial separation). Function (3) then becomes a fuzzy

† For ease of presentation, the source elements $\phi_1^e$, $\phi_2^e$, ... will always be ordered from left to right. In one dimension, pattern reversal actually does not pose any particular difficulty.

pattern probability density function:

$$P(\Phi)=\prod_{j}^{\alpha}\left[\left[\frac{2}{\alpha}\frac{1}{Q}\sum_{q}\left\{\sum_{l}W(l)\left[\prod_{s}^{\beta=2}C_s\exp\left(-\frac{(\phi_{j+L_{qs}}-\phi_s^e)^2}{2\sigma_s^2}\right)\right]\right\}\right.$$
$$\left.+\frac{(\alpha-2)}{\alpha}C_b\exp\left(-\frac{(\phi_j-\phi_b^e)^2}{2\sigma_b^2}\right)\right]\Bigg] \tag{4}$$

where $l\in L_{qs}\in l_{qs}\pm\Delta_{qs}$ and $\Delta_{qs}$ is the range of the voxels around the voxel $j\pm l_{qs}$ for which $W(l)$ is non-vanishing.

If there are $\beta>2$ spots making up the single species of fuzzy pattern, the probability density function (4) is further generalised:

$$P(\Phi)=\prod_{j}^{\alpha}\left[\left[\frac{\beta}{\alpha}\frac{1}{Q}\sum_{q}\left\{\sum_{l}W(l)\left[\prod_{s=1}^{\beta}C_s\exp\left(-\frac{(\phi_{j+L_{qs}}-\phi_s^e)^2}{2\sigma_s^2}\right)\right]\right\}\right.$$
$$\left.\times\left[+\frac{(\alpha-\beta)}{\alpha}C_b\exp\left(-\frac{(\phi_j-\phi_b^e)^2}{2\sigma_b^2}\right)\right]\right]\Bigg]. \tag{5}$$

A numerical example appears in appendix 2. Although the imaging results for simulation and experimental phantom data are carried out for the two-element pattern of function (4) only (as shown in §§ 4 and 5), the detailed description of the general function (5) in appendix 2 is intended to help in visualising the general multi-element pattern formalism.

The statistical nature of the measured data reflects properties of the imaging system and the sources. For radioisotopic measurement each data element $N_i$ (or projection $i$) obeys Possion statistics†. If all elements $\{N_i\}=\boldsymbol{N}(i=1,2,\ldots,\gamma)$ are statistically independent, then the conditional probability density function $P(\boldsymbol{N}|\Phi)$ of measured data $\boldsymbol{N}$ for a given set of voxel values $\Phi$ is:

$$P(\boldsymbol{N}|\Phi)=\prod_{i=1}^{\gamma}\left[\exp\left(-\sum_{j}R_{ij}\phi_j\right)\right]\left(\sum_{j}R_{ij}\phi_j\right)^{N_i}(N_i!)^{-1} \tag{6}$$

where $\sum_j R_{ij}\phi_j$ is the mean of the Poisson random variable $N_i$, $R_{ij}$ the probability of detecting photons from voxel $j$ in projection $i$ and $\gamma$ the number of data elements.

The values of $R_{ij}$ are in general a function of the voxel projection geometry, tissue attentuation and scattering. In tomographic imaging (e.g. CAT, PET and SPECT, etc), $R_{ij}$ is often calculated based on the geometric relations of voxels and projection rays (Brooks and Chiro 1976, Phelps 1977), and in many image restoration applications it is determined experimentally by imaging a point source, i.e. measuring the point spread function (PSF) (Andrews and Hunt 1977).

For all possible sets of voxel values $\{\Phi\}\in\Omega$ which would give rise to the projected data $\boldsymbol{N}$, the probability density function of $\boldsymbol{N}$ is:

$$P(\boldsymbol{N})=\int_{\Omega}P(\boldsymbol{N}|\Phi)P(\Phi)\,\mathrm{d}\Phi \tag{7}$$

where, as usual, $P(\boldsymbol{N})$ is the probability distribution in the data space, $P(\Phi)$ is the probability distribution in the source space (here an *a priori* probability distribution) and $P(\boldsymbol{N}|\Phi)$ is the conditional probability distribution in the data space for a given source field $\Phi$.

† Consideration of Gaussian data is analogous (Liang and Hart 1987).

### 3. The BIP formalism

The BIP formalism is directed at finding that source distribution $\Phi^* \in \Omega$ which is most likely to have given rise to the measured data, subject to the data vector $\boldsymbol{N}$, the matrix $\boldsymbol{R}$ and any *a priori* source distribution information which may be available (e.g. the $P(\Phi)$ function). Mathematically this objective is equivalent to maximising the conditional probability density function $P(\Phi|\boldsymbol{N})$ where from Bayes' law (Lambert 1968):

$$P(\Phi|\boldsymbol{N}) = P(\boldsymbol{N}|\Phi)P(\Phi)/P(\boldsymbol{N}). \tag{8}$$

Here, $P(\Phi|\boldsymbol{N})$ is the conditional probability distribution in the source space for a given data result $\boldsymbol{N}$.

The optimal solution $\Phi^*$ is then specified by maximising $P(\Phi|\boldsymbol{N})$, or equivalently $g(\Phi) \equiv \ln(P(\Phi|\boldsymbol{N}))$, algebraically by solving the system of equations:

$$\left.\frac{\partial g(\Phi)}{\partial \phi_k}\right|_{\Phi=\Phi^*} = 0. \tag{9}$$

Substituting for $P(\boldsymbol{N}|\Phi)$ of (6) and $P(\Phi)$ of (5) and neglecting $P(\boldsymbol{N})$ since it is not a function of $\Phi$, equation (9) becomes:

$$\sum_i R_{ik}\left(\frac{N_i}{\sum_j R_{ij}\phi_j^*} - 1\right) = \frac{Y_k(\phi_k^*) + Z_k(\phi_k^*)}{X_k(\phi_k^*)} \tag{10}$$

where $X_k(\phi_k^*)$, $Y_k(\phi_k^*)$ and $Z_k(\phi_k^*)$ are defined in appendix 3.

Equation (10) serves to determine that solution $\Phi^*$ which would most likely give rise to the measured data, subject to the probabilistic constraint of the estimated strength and spatial separation values of $\{\phi_s^e, \phi_b^e \text{ and } L_{qs}\}$. For a low-resolution imaging system (i.e. algebraically a $\boldsymbol{R}$ matrix reflecting an extended FWHM) and large dimension $\Phi$ and $\boldsymbol{N}$, an iterative approach to the solution $\Phi^*$ is preferred and the expectation maximisation (EM) technique (Dempster *et al* 1977), as well as others, can be employed to derive a BIP iterative algorithm.

In EM nomenclature, the 'complete data' variable $M_{ij}$ is now the number of photons emitted by voxel $j$ and contributing to projection $i$ and is related to the 'incomplete data' variable $N_i$ by $N_i = \sum_j M_{ij}$. $N_i$ and $M_{ij}$ correspond to $Y_i$ and $X_{ij}$ respectively in the notation of Lange and Carson (1984). Bayes' law for the 'complete data' $\boldsymbol{M}$ is: $P(\Phi|\boldsymbol{M}) = P(\boldsymbol{M}|\Phi)P(\Phi)/P(\boldsymbol{M})$. Since $M_{ij}$ is also a Poisson random variable (Feller 1968), function $P(\boldsymbol{M}|\Phi)$ has the similar form of function (6) (Liang and Hart 1987) and $P(\boldsymbol{M})$ is independent of $\Phi$ (analogous to function (7)) and will be omitted in the iterative solution for $\Phi^*$.

The ln Bayesian function of $P(\Phi|\boldsymbol{M})$ is:

$$\begin{aligned} f(\Phi) &\equiv \ln P(\Phi|\boldsymbol{M}) \\ &= \sum_{ij}[-R_{ij}\phi_j + M_{ij}\ln(R_{ij}\phi_j) - \ln(M_{ij}!)] + \sum_j \ln X_j(\phi_j) \end{aligned} \tag{11}$$

with $X_j(\phi_j)$ as defined in equation (A4).

The E-step (expectation) of the EM technique computes the conditional average value of

$$\begin{aligned} Q(\Phi|\Phi^{(n)}) &= E[f(\Phi)|\boldsymbol{N}, \Phi^{(n)}] \\ &= \sum_{ij}[-R_{ij}\phi_j + M_{ij}^{(n)}\ln(R_{ij}\phi_j) - \ln(M_{ij}^{(n)}!)] + \sum_j \ln X_j(\phi_j) \end{aligned} \tag{12}$$

where $\Phi^{(n)}$ represents the $n$th iterated result and $M_{ij}^{(n)}$, the conditional average value of the random variable $M_{ij}$ for the given projection $i(N_i)$, is (Shepp and Vardi 1982, Rao 1952):

$$M_{ij}^{(n)} = E[M_{ij}|N_i, \Phi^{(n)}] = N_i \frac{R_{ij}\phi_j^{(n)}}{\sum_k R_{ik}\phi_k^{(n)}}. \tag{13}$$

The M-step (maximisation) of the EM technique generates the $(n+1)$th iteration result $\Phi^{(n+1)}$ which maximises the function $Q(\Phi|\Phi^{(n)})$ (see equation (9)):

$$\sum_i \left( -R_{ik} + \frac{M_{ik}^{(n)}}{\phi_k^{(n+1)}} \right) = \frac{Y_k(\phi_k^{(n+1)}) + Z_k(\phi_k^{(n+1)})}{X_k(\phi_k^{(n+1)})} \tag{14}$$

with $X_k(\phi_k^{(n+1)})$, $Y_k(\phi_k^{(n+1)})$ and $Z_k(\phi_k^{(n+1)})$ defined as in equations (A4), etc.

The $\phi_k^{(n+1)}$ on the right-hand side of equation (14) can be approximated as $\phi_k^{(n+1)} = \phi_k^{(n)} + \tau d_k^{(n)}$; the BIP algorithm then becomes:

$$\phi_k^{(n+1)} = \phi_k^{(n)} \sum_i R_{ik} \left[ N_i \left( \sum_j R_{ij}\phi_j^{(n)} \right)^{-1} \right] \times \left[ \sum_i R_{ik} + \xi_k^{(n)} \left( \frac{Y_k^{(n)}(\phi_k^{(n)} + \tau d_k^{(n)}) + Z_k^{(n)}(\phi_k^{(n)} + \tau d_k^{(n)})}{X_k^{(n)}(\phi_k^{(n)} + \tau d_k^{(n)})} \right) \right]^{-1} \tag{15}$$

where $\tau$ and $d_k^{(n)}$ are approximated by:

$$\tau \simeq 1 \qquad \text{and} \qquad d_k^{(n)} = \phi_k^{(n)} - \phi_k^{(n-1)} \tag{16}$$

and $\xi_k^{(n)}$ is an adjustable parameter introduced for numerical implementation of equation (15) and is chosen to be a monotonic sigmoidal function of the iterative index $n$ of the form:

$$\xi_k^{(n)} = \frac{An^\nu}{B + n^\nu} \sum_i R_{ik} \tag{17}$$

which gradually modifies the relative importance of the *a priori* source information function $P(\Phi)$. Such gradual modification is necessary because for early iterations $\phi_j^{(n)} \simeq N_j$ ($n$ small) and $P(\Phi)$ is then approximately equal to 0. Therefore successive iterative source estimates would diverge widely. By only gradually imposing the full restrictiveness of a Bayesian *a priori* constraint, the maximum likelihood (ML) algorithm of Lange and Carson (1984) (i.e. setting $\xi_k^{(n)} = 0$ in equation (15)) dominates initially and the iterative solution converges to the neighbourhood of the optimal solution where $P(\Phi)$ no longer vanishes. BIP then serves to generate the overall maximum likelihood source estimate reflecting both the imaging data and the *a priori* source information. The parameters $\nu$, $A$ and $B$ are numerical constants. The function $X_k^{(n)}(\phi_k^{(n)} + \tau d_k^{(n)})$ is again specified by equation (A4) (i.e. equation (A5) in appendix 3) and similarly for $Y_k^{(n)}(\phi_k^{(n)} + \tau d_k^{(n)})$ and $Z_k^{(n)}(\phi_k^{(n)} + \tau d_k^{(n)})$.

## 4. Simulation results

The one-dimensional source distribution chosen $\{S_j\}$, as shown in figure 1 by the full curve, consists of two point sources of 50 units strength and of 8 voxels separation ($L = 8$), superimposed on a uniform background of 10 units strength ($S_{11} = S_{19} = 50 + 10 = 60$). The one-dimensional noise-free or ideal data denoted by the dotted curve in figure 1 were calculated from $\sum_j R_{ij}S_j$. The non-vanishing elements of the source

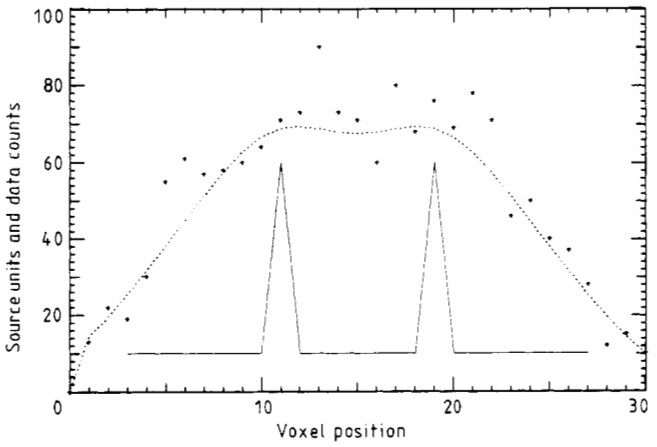


**Figure 1.** One-dimensional source (full curve) and data distributions. The dotted curve represents the ideal data distribution. The asterisks represent computer randomised data (Poisson).

distribution range from voxel position 3 to 27 and the ideal data distribution extends two voxels beyond the graph on both sides (i.e. five voxels beyond the non-vanishing elements of the source distribution). The $R_{ij}$ was chosen as:

$$R_{ij} = 0.5 \exp\{-(\ln 2)[(i-j)/T]^2\} \qquad i = -2, -1, 0, 1, \ldots, 32$$
$$j = 3, 4, 5, 6, \ldots, 27 \qquad T = 4 \tag{18}$$

defining a matrix of $35 \times 25$ elements.

A single set of Poisson randomised data (denoted by asterisks in figure 1) was obtained from the ideal data using a standard Poisson random number generator (Carnahan *et al* 1978).

For the two-point source pattern considered, the fuzzy pattern function (4) applies. The terms defined by equation (A5), etc, in the denominator of the BIP algorithm equation (15) have the form:

$$X_k^{(n)}(\ldots) = \eta_k^{(n)} \exp(-U_b(\phi_k^{(n)})) + \left( \sum_{l \in l_1 \pm \Delta} W(l) \exp(-U_s(\phi_k^{(n)})) + \sum_{l \in l_1 \pm \Delta} W(l) \exp(-V_s(\phi_k^{(n)})) \right)$$

$$Y_k^{(n)}(\ldots) = \eta_k^{(n)} \frac{(\phi_k^{(n)} + \tau d_k^{(n)} - \phi_b^e)}{\sigma_b^2} \exp(-U_b(\phi_k^{(n)})) \tag{19}$$

and

$$Z_k^{(n)}(\ldots) = \frac{(\phi_k^{(n)} + \tau d_k^{(n)} - \phi_1^e)}{\sigma_1^2} \sum_{l \in l_1 \pm \Delta} W(l) \exp(-V_s(\phi_k^{(n)})) + \frac{(\phi_k^{(n)} + \tau d_k^{(n)} - \phi_2^e)}{\sigma_2^2} \sum_{l \in l_1 \pm \Delta} W(l) \exp(-U_s(\phi_k^{(n)}))$$

where $V_s(\phi_k^{(n)})$, $U_b(\phi_k^{(n)})$ and $U_s(\phi_k^{(n)})$ are detailed by equation (A6), etc, in appendix 3.

The $\phi_1^e$ and $\phi_2^e$ in equation (19) are the estimated strengths of the assumed two spots respectively and $l \in l_1 \pm \Delta$ corresponds to $L_{qs}$ defined as in function (4). The

adjustable parameter $\eta_k^{(n)}$ is a function of $C_s$ and $C_b$ and was chosen to have the functional form:

$$\eta_k^{(n)} = \frac{A_0 n^\nu}{B_0 + n^\nu}. \tag{20}$$

The *a priori* estimate $\phi_s^e$ and the variance parameter $\sigma_s$ are also represented as functions of the iterative index $n$:

$$\phi_s^e \to \phi_b^e + \frac{\phi_s^e - \phi_b^e}{\sqrt{D}}\sqrt{n} \qquad \sigma_s^2 = \phi_s^e/\sqrt{n} \qquad s = 1, 2 \tag{21}$$

for numerical implementation of equation (15), where $D$ is the total number of iterations. The spatial weighting function $W(l)$ was assumed to be of the form:

$$W(l) = \frac{\Gamma^2}{\Gamma^2 + (l - l_1)^2} \tag{22}$$

with $l \in l_1 \pm \Delta$ as employed in equation (19), etc. In implementing the BIP algorithm expressed by equation (15), the numerical constants in equations (17), (20) and (22) were chosen as:

$$A_0 \simeq 1 \qquad A \simeq 0.1 \qquad B_0 \simeq B \simeq 100 \qquad \Gamma = 1 \qquad \nu = 2.$$

Figure 2 shows the ideal data results of the ML algorithm (Lange and Carson 1984) after 100 iterations and of the BIP algorithm after only 50 iterations†. In using the BIP algorithm, it was assumed that $\phi_1^e = 55$, $\phi_2^e = 65$ units rather than the actual values of 60 units each. The estimated separation $l_1$ of the two spots was seven voxels rather than the actual separation of eight voxels. The weighting function $W(l)$ assumes a range of $\Delta = 2$, i.e. $l = 5, 6, 7, 8, 9$, and for this set of calclations $W(l) = 0.2, 0.5, 1, 0.5, 0.2$. Note that in spite of the imperfect *a priori* assumption of a separation of seven

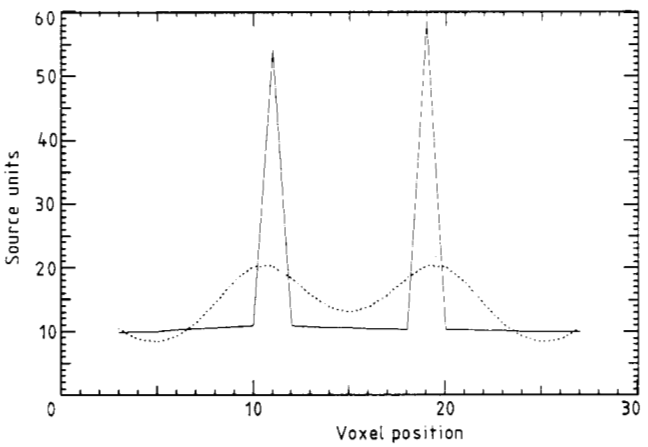


**Figure 2.** Results using algorithm BIP at 50 iterations (full curve) and algorithm ML at 100 iterations (dotted curve) on ideal data.

† Note that the presence of the *a priori* probability constraint $P(\Phi)$ tends to suppress the iterative divergences common to most other methods. The results after 50 iterations using the BIP algorithm were generally indistinguishable from those after 100 iterations.

voxels and the inexact amplitudes of 55 and 65 units, the BIP maximum likelihood solution matches the chosen source distribution $\{S_j\}$ very well. For assumed separations of $l_1 = 6$ or 10, the BIP solutions remained excellent although the peaks were slightly broader. However, for assumed separations of 5 or 11 corresponding for $\Delta = 2$ to the *a priori* probabilities $W(l = 5, 11) = 0$, the BIP solutions were still slightly superior to the ML result. This unexpected residual small improvement may be attributable to the relatively accurate amplitude estimates.

The results of both algorithms ML and BIP applied to the single set of random data of figure 1 after 100 and 50 iterations respectively are shown in figure 3, where for the BIP algorithm $\phi_1^e = 55$, $\phi_2^e = 65$ and $l_1 = 7$ were assumed, and $l \in l_1 \pm \Delta = 7 \pm 2$ voxels (i.e. the same *a priori* $P(\Phi)$ distribution used for the ideal data case of figure 2). Since the resulting maxima (one set, e.g. as shown in figure 3) fall on or shift around the positions (11, 19) of the actual two-point sources for different sets of random data, the algorithms were then applied to ten sets of random data and the averages of corresponding results are shown in figure 4.

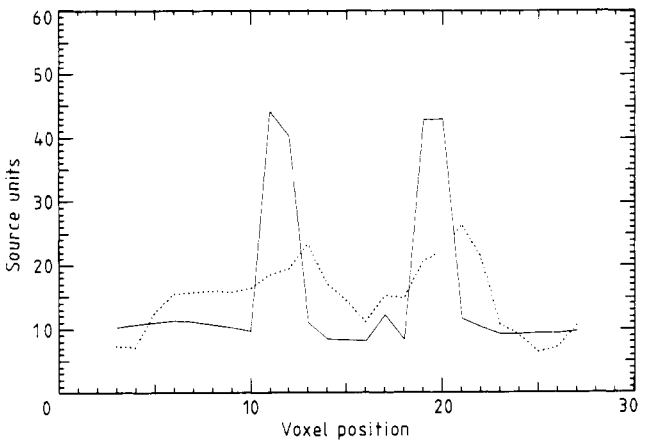


**Figure 3.** Results using algorithm BIP at 50 iterations (full curve) and algorithm ML at 100 iterations (dotted curve) on a single set of computer randomised data.

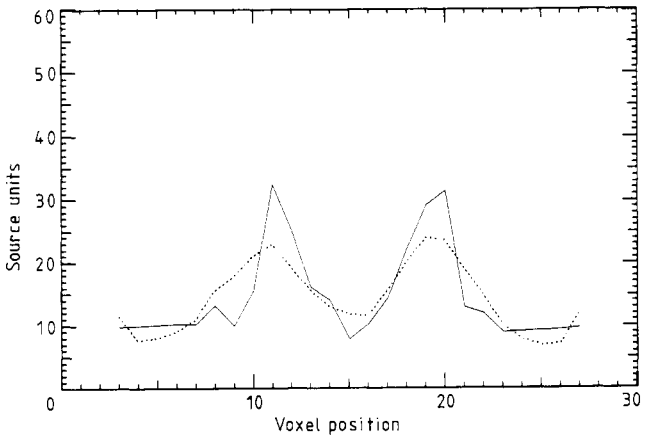


**Figure 4.** Comparison of BIP at 50 iterations (full curve) with ML at 100 iterations (dotted curve) for ten sets of computer randomised data.

It is easily seen that the BIP algorithm resulted in substantial improvement over the ML algorithm. For assumed values of $l_1 = 6$ or 10 and $\Delta = 2$, the BIP solutions remain superior to the ML result. For $l_1 = 5$ or 11, the BIP results on average were comparable with that of ML.

To quantitate the deviation of the resulting source distribution $\{\phi_j^{(n)}\}$ from the chosen source distribution $\{S_j\}$, a test function was introduced,

$$\delta = \sum_j (\bar{\phi}_j^{(n)} - \bar{S}_j)^2 / \bar{S}_j \tag{23}$$

with weighted averages of $\phi_j^{(n)}$ and $S_j$ defined as

$$\bar{S}_j = \sum_r R_{rj} S_r \left( \sum_r R_{rj} \right)^{-1} \qquad \bar{\phi}_j^{(n)} = \sum_r R_{rj} \phi_r^{(n)} \left( \sum_r R_{rj} \right)^{-1} \tag{24}$$

where $r \in j \pm \Lambda$, $\Lambda \simeq \frac{1}{2}T$ specifies the maximal neighborhood of the PSF. Function (23) gives a more accurate indication than the standard test function (Wine 1964):

$$\delta_0 = \sum_j (\phi_j^{(n)} - S_j)^2 / S_j. \tag{25}$$

For example, $\delta = 11.60$ and 31.14 for the results of the BIP and ML algorithms respectively shown in figure 3, in which $\Lambda = 2$ was chosen. Using function (25) $\delta_0 = 211.4$ and 170.2 respectively actually reversing the subjectively obvious order of agreement. The inappropriateness of the test function $\delta_0$ for this class of applications arises from its relative inflexibility in evaluating spatial positioning errors. A single voxel shift in the calculated position of a point source compared with its true location results in a very large change in value of $\delta_0$ and implies a grossly incorrect result. By reflecting the extent of spatial shifts quantitatively, the test function $\delta$ somewhat more realistically assesses the relative accuracy of image processing results.

The amplitude uncertainties in the *a priori* fuzzy pattern probability distribution function (5), $P(\Phi)$, are reflected by the $\sigma_s$ parameters, and the spatial uncertainties by the $W(l)$ weighting functions. Numerous calculations using a range of $\sigma_s$ values and $W(l)$ functions indicate that:

(a) very accurate results are obtained with BIP when the actual source distribution lies within the central range of the multidimensional $P(\Phi)$ distribution even for extremely noisy data. The results are generally significantly better than those obtained using the ML algorithm;

(b) if the source distribution lies completely outside the range of the *a priori* information function $P(\Phi)$, the BIP results are no better and often can be worse than the ML results.

## 5. Experimental results

A point source of $^{57}$Co was imaged using a Picker Dyna Camera model No 4 and the output was, for simplicity, arranged as a $32 \times 32$ matrix (i.e. a two-dimensional point source response function (PSF), $F_{ij}$), from which a line source response function (LSF), $L_i$, could be determined:

$$L_i = \sum_j F_{ij} \tag{26}$$

and is shown in figure 5 by a full curve in which the centre value is normalised to 40.

Since the imaging system was approximately spatially invariant, a matrix representation, $R_{ij}^L$, of the one-dimensional LSF could be formed from $\{L_i\}$ (appendix 1 of Andrews and Hunt 1977).

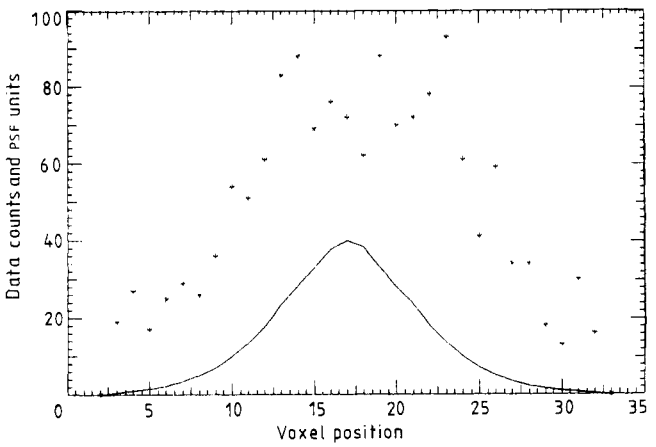


**Figure 5.** The phantom imaging data from a centre column of the two-dimensional data matrix (asterisks) and the line source response function (full curve) for $^{57}$Co.

A simple one-dimensional equivalent phantom was prepared by threading two parallel catheters containing a solution of $^{57}$Co through a stainless steel screen. Two-dimensional data were obtained by imaging the phantom at the same depth as the point source and were arranged as a $32 \times 32$ matrix, with the two lines of tubing oriented in the row direction. Column 16 of the data matrix is shown in figure 5 by asterisks and was used as the one-dimensional imaging data $\{N_i\}$. Neglecting the effect of the finite length of the tubing, $\{N_i\}$ can be viewed as imaging data from a double element source distribution (the projection of the two line sources along the parallel direction superimposed on a uniform background) in a one-dimensional geometry. Algebraically then:

$$N_i = \sum_j R_{ij}^L S_j \tag{27}$$

where $\{N_i\}$ is column 16 of the data matrix, $R_{ij}^L$ is the convolution matrix generated by the LSF and $\{S_j\}$ is the equivalent one-dimensional phantom.

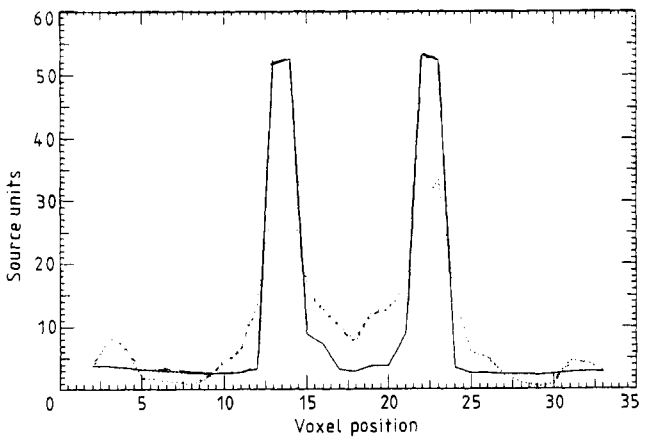


**Figure 6.** Comparison of BIP at 50 iterations (full curve) with ML at 100 iterations (dotted curve) for the phantom imaging data shown in figure 5.

The parameters for $W(l)$ of equation (22) were $l_1 = 10$, $\Delta = 3$ and $\Gamma = 1.5$ (the actual separation of the lines of tubing was approximately 7). To obtain the rough estimates of the amplitude parameters $\phi_b^e = 3$ and $\phi_1^e = \phi_2^e = 60$ assumed in the *a priori* information, $P(\Phi)$, experimental data obtained with and without the phantom were analysed making use of the factor $\sum_{ij} F_{ij}$. This was indicated since neither the absolute sensitivity of the equipment nor the absolute activity of the phantom were known. Figure 6 demonstrates the superior result achieved with the BIP algorithm (full curve) by 50 iterations compared with the result obtained using the ML algorithm (dotted curve) after 100 iterations. Similarly improved results were obtained with BIP for all of the centrally located data columns.

## 6. Discussion

As is well known (Hunt 1977, Herman *et al* 1979, Liang and Hart 1987), the BIP formalism has the advantage of considering *both* the statistical nature of the measured data and any *a priori* source distribution information available.

As indicated in our previous work (Liang and Hart 1987, Hart and Liang 1987), the information incorporated in the *a priori* probability density function $P(\Phi)$ need not be restricted to generic, essentially qualitative characterisations of the source field $\Phi$. In this paper a formalism is outlined and tested for image processing in which the *a priori* information includes more restrictive amplitude and pattern characteristics of $\Phi$.

The *a priori* probability density functions $P(\Phi)$ generally reflect multidimensional source distributions. For two- and three-dimensional geometries, uncertainties in the orientations of the anticipated patterns substantially increase the software and hardware computational requirements. Multidimensional studies with computer generated data and experimentally measured data from phantoms are in progress.

In some realistic medical situations, the *a priori* information may be quite complex even to the extent of reflecting subjective qualitative evaluations based upon extensive clinical experience. Characterisation of the useful *a priori* probability function $P(\Phi)$ may then involve relatively sophisticated judgments. It should be emphasised that, as indicated at the end of § 4, any improvement in image processing is crucially dependent upon the validity of this *a priori* information. If the true source distribution is outside the range of $P(\Phi)$, BIP results can be not only quantitatively inferior to those of ML, but artefacts may even be possible.

Note that $P(\Phi)$ is only approximately given by $\prod_j P_j(\phi_j)$ since the possible existence of a pattern implies of course a dependence between appropriate voxel values. In one-dimensional geometries, the effect of this dependence may not be completely trivial. In two and three dimensions, the inclusion of a single voxel $j$ in a pattern does not specify the pattern orientation and the statistical dependence of regional voxel values arising from the possible presence of a pattern is correspondingly weaker. The probability of a given sequence of voxels corresponding in amplitude and spatial orientation to a fuzzy pattern on a random basis is, however, likely to be small and the approximation $P(\Phi) = \prod_j P_j(\phi_j)$ which actually tends to understate the probability of a pattern being present does not seem to compromise the calculated results.

Function (2) can be generalised to characterise $\beta' > 1$ estimated source strengths $\{\phi_s^e\}$, $s = 1, 2, \ldots, \beta'$, and may be applicable, for example, in cardiac imaging as discussed in § 1. The application of the generalised fuzzy pattern function (5) is also mentioned in § 1, where the strengths and spacing information of the fuzzy pattern

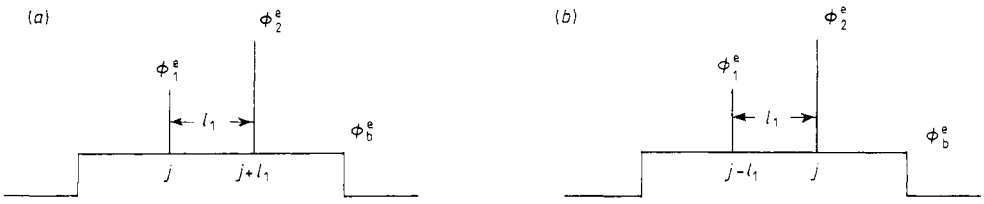


**Figure 7.** Two possible configurations of a two-element pattern.

elements can be obtained, in a simple way, by use of a standard non-Bayesian algorithm (e.g. ML) before using the BIP algorithm.

The statistical nature of the measured data is incorporated accurately into the conditional probability density function $P(\boldsymbol{N}|\Phi)$ when all elements of the measured data are statistically independent. For statistically correlated data elements, $P(\boldsymbol{N}|\Phi)$ is not in general easily formulated. If the statistical correlation between neighbouring source elements is Gaussian, then quadratic optimisation image processing may be applicable (Liang and Hart 1987).

## Appendix 1

As shown in figures 7($a$) and 7($b$), there are two possible configurations for which either element of a two-element pattern occupies the voxel $j$. For configuration 7($a$), the amplitude probability distribution for voxel $j$ is $C_1 \exp[-(\phi_j - \phi_1^e)^2/(2\sigma_1^2)]$; the amplitude probability distribution for voxel $j+l_1$ is $C_2 \exp[-(\phi_{j+l_1} - \phi_2^e)^2/(2\sigma_2^2)]$. The two-parameter amplitude probability distribution $P_j(\phi_j, \phi_{j+l_1})$ for configuration 7($a$) is then $C_1 C_2 \exp[-(\phi_j - \phi_1^e)/(2\sigma_1^2) - (\phi_{j+l_1} - \phi_2^e)^2/(2\sigma_2^2)]$.

Assuming that the probabilities of the two configurations 7($a$) and 7($b$) are equal and viewing $\phi_{j\pm l_1}$ now as fixed parameters, the amplitude probability distribution for a voxel $j$ included in the two-element pattern is:

$$\tfrac{1}{2}C_1C_2\{\exp[-(\phi_{j-l_1} - \phi_1^e)^2/(2\sigma_1^2) - (\phi_j - \phi_2^e)^2/(2\sigma_2^2)] + \exp[-(\phi_j - \phi_1^e)^2/(2\sigma_1^2) - (\phi_{j+l_1} - \phi_2^e)^2/(2\sigma_2^2)]\}.$$

Taking into account the background probability distribution, function (3) follows.

## Appendix 2

Assume as in figure 8 a three-element pattern. If

$$W(\Delta_{qs}) = \begin{cases} 0.25 & \Delta_{qs} = -1 \\ 0.50 & \Delta_{qs} = 0 \\ 0.25 & \Delta_{qs} = 1 \end{cases} \tag{A1}$$

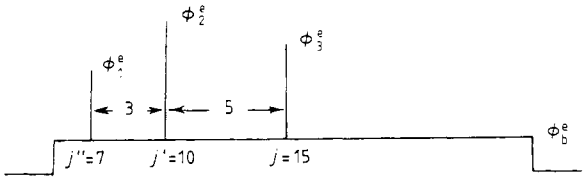


**Figure 8.** Three-element pattern; configuration $q=1$ with $\phi_3^e$ corresponding to $\phi_{15}$.

the *a priori* probability of configuration $q=1$ with $\phi_3^e$ corresponding as in figure 8 to the voxel of the arbitrarily chosen $j=15$ is proportional to

$$\begin{aligned}P_{q=1}(\phi_{15}) \simeq & \exp[-(\phi_{15}-\phi_3^e)^2/(2\sigma_3^2)] \\ & \times\{0.125\exp[-(\phi_6-\phi_1^e)^2/(2\sigma_1^2)-(\phi_{10}-\phi_2^e)^2/(2\sigma_2^2)] \\ & +0.25\exp[-(\phi_7-\phi_1^e)^2/(2\sigma_1^2)-(\phi_{10}-\phi_2^e)^2/(2\sigma_2^2)] \\ & +0.125\exp[-(\phi_8-\phi_1^e)^2/(2\sigma_1^2)-(\phi_{10}-\phi_2^e)^2/(2\sigma_2^2)] \\ & +0.125\exp[-(\phi_7-\phi_1^e)^2/(2\sigma_1^2)-(\phi_9-\phi_2^e)^2/(2\sigma_2^2)] \\ & +0.125\exp[-(\phi_7-\phi_1^e)^2/(2\sigma_1^2)-(\phi_{11}-\phi_2^e)^2/(2\sigma_2^2)] \\ & +0.0625\exp[-(\phi_6-\phi_1^e)^2/(2\sigma_1^2)-(\phi_9-\phi_2^e)^2/(2\sigma_2^2)] \\ & +0.0625\exp[-(\phi_6-\phi_1^e)^2/(2\sigma_1^2)-(\phi_{11}-\phi_2^e)^2/(2\sigma_2^2)] \\ & +0.0625\exp[-(\phi_8-\phi_1^e)^2/(2\sigma_1^2)-(\phi_9-\phi_2^e)^2/(2\sigma_2^2)] \\ & +0.0625\exp[-(\phi_8-\phi_1^e)^2/(2\sigma_1^2)-(\phi_{11}-\phi_2^e)^2/(2\sigma_2^2)]\}. \end{aligned} \tag{A2}$$

Configuration $q=2$ corresponds to $\phi_2^e=\phi_{15}$, and configuration $q=3$ corresponds to $\phi_1^e=\phi_{15}$ with similar expressions for $P_{q=2}(\phi_{15})$ and $P_{q=3}(\phi_{15})$.

Finally the likelihood of voxel $j=15$ being included in the fuzzy pattern is:

$$P(\phi_{15}) \simeq P_{q=1}(\phi_{15})+P_{q=2}(\phi_{15})+P_{q=3}(\phi_{15}). \tag{A3}$$

## Appendix 3

The $X_k(\phi_k^*)$, $Y_k(\phi_k^*)$ and $Z_k(\phi_k^*)$ in equation (10) have the forms:

$$\begin{aligned} X_k(\phi_k^*) &= \sum_q \left\{ \sum_l W(l)\left[ \prod_s C_s \exp\left( -\frac{(\phi^*_{k+L_{qs}}-\phi_s^e)^2}{2\sigma_s^2}\right)\right]\right\} + C_b \exp\left( -\frac{(\phi_k^*-\phi_b^e)^2}{2\sigma_b^2}\right) \\ Y_k(\phi_k^*) &= \frac{(\phi_k^*-\phi_b^e)}{\sigma_b^2} C_b \exp\left( -\frac{(\phi_k^*-\phi_b^e)^2}{2\sigma_b^2}\right) \\ Z_k(\phi_k^*) &= \sum_q \left\{ \frac{(\phi_k^*-\phi_q^e)}{\sigma_q^2} \sum_l W(l)\left[ \prod_s C_s \exp\left( -\frac{(\phi^*_{k+L_{qs}}-\phi_s^e)^2}{2\sigma_s^2}\right)\right]\right\} \end{aligned} \tag{A4}$$

with $\alpha$, $\beta$ and $Q$ incorporated into $C_s$ and $C_b$. Similarly,

$$\begin{aligned} & X_k^{(n)}(\phi_k^{(n)}+\tau d_k^{(n)}) \\ & \quad = \sum_q \left\{ \sum_l W(l)\left[ \prod_s C_s \exp\left( -\frac{(\phi^{(n)}_{k+L_{qs}}+\tau d^{(n)}_{k+L_{qs}}-\phi_s^e)^2}{2\sigma_s^2}\right)\right]\right\} \\ & \quad + C_b \exp\left( -\frac{(\phi_k^{(n)}+\tau d_k^{(n)}-\phi_b^e)^2}{2\sigma_b^2}\right). \end{aligned} \tag{A5}$$

The $V_s(\phi_k^{(n)})$, $U_b(\phi_k^{(n)})$ and $U_s(\phi_k^{(n)})$ defined in equation (19) are expressed as:

$$\begin{aligned} V_s(\phi_k^{(n)}) &= \frac{(\phi_k^{(n)}+\tau d_k^{(n)}-\phi_1^e)^2}{2\sigma_1^2}+\frac{(\phi_{k+l}^{(n)}+\tau d_{k+l}^{(n)}-\phi_2^e)^2}{2\sigma_2^2} \\ U_b(\phi_k^{(n)}) &= \frac{(\phi_k^{(n)}+\tau d_k^{(n)}-\phi_b^e)^2}{2\sigma_b^e} \end{aligned} \tag{A6}$$

and

$$U_s(\phi_k^{(n)}) = \frac{(\phi_{k-l}^{(n)}+\tau d_{k-l}^{(n)}-\phi_1^e)^2}{2\sigma_1^2}+\frac{(\phi_k^{(n)}+\tau d_k^{(n)}-\phi_2^e)^2}{2\sigma_2^2}.$$

## Résumé

Traitement d'images Bayesien des données de distributions de sources contraintes. Contraintes du modèle flou.

Les auteurs ont développé des fonctions de densité de probabilité *à priori* caractérisant des modèles qui sont imprécis spatialement et en amplitude (modèles flous) et dont on suppose la présence dans le champ d'une source radioactive pour une utilisation par un traitement d'image Bayesien (BIP). En utilisant la technique de maximisation de l'espérance (EM) mathématique proposée par Dempster *et al*, ils ont obtenu des algorithmes itératifs de traitement d'image. Des algorithmes BIP et des algorithmes standards non BIP ont été appliqués a des données générées expérimentales obtenues au moyen de fantômes radioactifs. Des résultats améliorés ont étés obtenus avec BIP.



## Zusammenfassung

Bayes-Bildverarbeitung von Daten zwangsverteilter Quellen—Zwangsbedingungen unscharfer Muster.

Zur Verwendung in der Bayes-Bildverarbeitung (BIP) wurden Anfangswahrscheinlichkeitsdichtefunktionen entwickelt. Diese charakterisieren räumlich und bzgl. ihrer Amplitude unscharfe Muster, deren Vorhandensein in Quellfeldern von Radioisotopen bereits vorhergesagt wurden. Entsprechende iterative Abbildungsalgorithmen wurden hergeleitet mit Hilfe des Erwartungs-Maximierungs-(EM)-Verfahrens von Dempster *et al.* BIP- und nicht-BIP-Standardalgorithmen wurden angewendet auf Computer-erzeugte und experimentell gewonnene Abbildungsdaten eines Radioisotopen-Phantoms. Mit dem BIP-Verfahren erhält man verbesserte Ergebnisse.


## References


Andrews H C and Hunt B R 1977 *Digital Image Restoration* (Englewood Cliffs, NJ: Prentice-Hall)
Brooks R A and Chiro G D 1976 *Phys. Med. Biol.* **21** 689
Carnahan B, Luther H A and Wilkes J D 1978 *Applied Numerical Methods* (New York: Wiley)
Dempster A P, Laird N M and Rubis D B 1977 *J. R. Stat. Soc.* B **39** 1
Feller W 1968 *An Introduction to Probability Theory and its Applications* (New York: Wiley)
Hart H and Liang Z 1987 *Bull. Math. Biol.* **49** 75
Herman G, Hurwitz H, Lent A and Hsi-Ping Lung 1979 *Inform. Control* **42** 60
Hunt B 1977 *IEEE Trans. Comput.* **C-3** 219
Lambert R F 1968 *An Introduction to Random Signals and Communication Theory* (Scranton, PA: International Textbook Company)
Lange K and Carson R 1984 *J. Comput. Assist. Tomogr.* **8** 306
Liang Z and Hart H 1987 *Bull. Math. Biol.* **49** 51
Phelps M E 1977 *Sem. Nucl. Med.* **7** 337
Rao C 1952 *Advanced Statistical Methods in Biometric Research* (New York: Wiley)
Shepp L A and Vardi Y 1982 *IEEE Trans. Med. Imag.* **MI-1** 113
Wine R L 1964 *Statistics for Scientists and Engineers* (Englewood Cliffs, NJ: Prentice-Hall)